

# Ancient human mtDNA genotypes from England reveal lost variation over the last millennium

A. L. Töpf<sup>1</sup>, M. T. P. Gilbert<sup>2,3</sup>, R. C. Fleischer<sup>4</sup> and A. R. Hoelzel<sup>1,\*</sup>

<sup>1</sup>School of Biological and Biomedical Sciences, University of Durham, Durham DH1 3LE, UK

<sup>2</sup>Ancient Biomolecules Centre, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

<sup>3</sup>Ancient DNA and Evolution, Niels Bohr Institute, University of Copenhagen, Copenhagen 2100, Denmark

<sup>4</sup>NMNH—Genetics Program, Smithsonian Institution, Washington, DC 20008, USA

\*Author for correspondence (a.r.hoelzel@durham.ac.uk).

We analysed the historical genetic diversity of human populations in Europe at the mtDNA control region for 48 ancient Britons who lived between ca AD 300 and 1000, and compared these with 6320 modern mtDNA genotypes from England and across Europe and the Middle East. We found that the historical sample shows greater genetic diversity than for modern England and other modern populations, indicating the loss of diversity over the last millennium. The pattern of haplotypic diversity was clearly European in the ancient sample, representing each of the modern haplogroups. There was also increased representation of one of the ancient haplotypes in modern populations. We consider these results in the context of possible selection or stochastic processes.

**Keywords:** ancient DNA; human; mtDNA; genetic diversity

## 1. INTRODUCTION

Since the appearance of anatomically modern humans in Europe in the Late Upper Palaeolithic (approx. 40 000 years BP), there have been demographic changes, which can be interpreted from patterns and levels of genetic diversity. This includes an apparent loss of diversity in Europe during the Holocene (e.g. Marth *et al.* 2004), signatures of population expansions (e.g. Comas *et al.* 1997) and Neolithic data suggesting primarily Palaeolithic ancestry in modern Europe (Haak *et al.* 2005). More recent historical events, especially widespread epidemics during the Middle ages, provided the potential for the stochastic loss of diversity (or selection). For example, it is known that while some entire families and even villages were wiped out by the plague, others were immune or resistant to it (Clifford 1989). In this study (after Töpf *et al.* 2006), we use archaeological material to assess historical matrilineal diversity in Britain around the time of the Roman and Saxon invasions, and find direct evidence for the loss of diversity over the last millennium.

## 2. MATERIAL AND METHODS

### (a) Samples

Human skeletal remains exhumed from five archaeological sites dated from the fourth to eleventh centuries were studied (see Töpf *et al.* (2006) for details and mtDNA HVS-I sequences; table 1). The comparative dataset from modern European and Middle Eastern populations were taken from GenBank, the HVR database (corrected version) and MITOMAP ( $N=6320$ ). In particular, modern sequences from England were from Piercy *et al.* (1993), Miller *et al.* (1996), Richards *et al.* (1996) and Helgason *et al.* (2001).

Thorough precautions against the inclusion of contaminant or false sequences were undertaken (see Töpf *et al.* 2006); however, the possible generation of false diversity from chimeras during amplification is a special consideration for this study. No evidence was detected from sequencing a subset of clones (Töpf *et al.* 2006), but as an extra precaution we used split decomposition analysis (implemented in SPLITSTREE v. 4; Huson 1998) to identify network-like within a phylogenetic tree as a way to look for possible chimeras (cf. Holmes *et al.* 1999). Chimeric sequences or post-mortem changes are unlikely to replicate full haplotypes already identified in the modern database (though individual sites may correspond, see Gilbert *et al.* (2003)); so as a first test, diversity levels were recalculated omitting individuals whose haplotypes were both unique to the ancient sample and involved in any level of networking within the split decomposition tree (test 1 in table 1). More stringent tests omitted all haplotypes involved in networking (test 2) or all haplotypes unique to the ancient sample (test 3).

### (b) Summary statistics

Haplotype and nucleotide diversity were estimated taking into account the differences in sample size among the populations. Haplotype diversity was assessed as Nei's unbiased gene diversity ( $h$ ), the Shannon index and its standardized version (where the index is divided by the maximum possible value for a given sample size; Magurran 1988), and 'haplotype richness', denoted as  $r(g)$  (based on Petit *et al.* 1998; implemented in Rarefactor, <http://www2.biology.ualberta.ca/jbrzusto/rarefact.php>). The proportion of pairwise differences ( $\pi$ ) between all pairs of sequences in the sample was calculated, assuming the Tamura–Nei model and a gamma distribution with  $\alpha=0.26$ . The level of unbiased gene diversity ( $h$ ) was statistically compared among populations using the conservative method proposed by Thomas *et al.* (2002), whereby both bootstrap resampling (10 000 iterations) and standard  $z$ -tests are implemented, and the higher of the two  $p$ -values used.

### (c) Phylogenetic analysis

Sequences were aligned using the SEQUENCHER v. 3.0 (Gene Codes Corporation, USA) and polymorphic positions identified using MEGA v. 2.1. Reduced median networks (Bandelt *et al.* 1995) were constructed using NETWORK v. 2.0. To maximize the resolution, all of the segregating sites found along the 207 bp (including deletions) were used for the analyses. A mismatch distribution was computed using ARLEQUIN and  $\tau$  calculated from the distribution using the same program.

## 3. RESULTS AND DISCUSSION

From 48 individuals, 36 haplotypes were authenticated and included (accession numbers DQ191964–DQ192011; Töpf *et al.* 2006). The reduced median network (RMN) for the ancient populations (figure 1) shows no indication of strong lineage sorting, as most of the major Eurasian haplogroups and sub-haplogroups can be identified in the network (except for H, preHV, J, U3 and U4).

Figure 2 shows the mismatch distributions for the ancient samples (all combined) and the samples from modern England. The crest of the distribution for the ancient population ( $\tau=4.78$ ) is shifted to the right compared with modern England ( $\tau=2.44$ ), suggesting an earlier population expansion for the ancient sample. This effect would be expected following a loss of diversity.

Estimates of genetic diversity (corrected for differences in sample size) were computed for historical and modern samples. Counter-intuitively, the genetic diversity of the ancient sample was consistently higher (across all measures) than the diversity observed in

Table 1. Mitochondrial DNA diversity values. (Combined Modern: England, Wales, northern Germany, Denmark and Norway. Approximate dates: Leicester, fourth century. Norton, fifth to sixth century; Buckland, sixth century, Lavington, sixth century and Norwich, tenth century. Number of original samples is given parenthetically, among which 'N' amplifications were successful. Number of haplotypes ( $k$ ), gene diversity  $\pm$  s.d. ( $h \pm$  s.d.), Shannon diversity index ( $H_s$ ), standardized Shannon index ( $H'_s = H_s/H_{s\max}$ ), haplotypic richness ( $r(g)$ )—based on a sample  $g=18$ , and % nucleotide diversity  $\pm$  s.d. ( $\pi \pm$  s.d.). Data in plain text are for subsets of ancient Britain and modern England.)

population	$N$	$k$	$k/N$	$h \pm$ s.d.	$H_s$	$H'_s$	$r_{(18)}$	$\pi(\%) \pm$ s.d.
<b>ancient Britain</b>	<b>48</b>	<b>36</b>	<b>0.75</b>	<b>0.99 <math>\pm</math> 0.07</b>	<b>−3.52</b>	<b>0.90</b>	<b>16.18</b>	<b>2.68 <math>\pm</math> 1.30</b>
Leicester (15)	6	6	1.00	1.00 $\pm$ 0.09	−1.79	1.00	n.a.	2.94 $\pm$ 1.63
Norton (23)	12	11	0.92	0.98 $\pm$ 0.04	−2.37	0.95	n.a.	2.75 $\pm$ 1.42
Buckland (31)	7	11	0.71	0.91 $\pm$ 0.01	−1.75	0.90	n.a.	2.55 $\pm$ 1.40
Lavington (28)	6	5	0.83	0.93 $\pm$ 0.01	−1.56	0.87	n.a.	1.15 $\pm$ 0.72
Norwich (59)	17	15	0.88	0.98 $\pm$ 0.02	−2.67	0.94	16.00	2.80 $\pm$ 1.41
test 1	44	33	0.75	0.99 $\pm$ 0.01	−3.40	0.90	15.92	2.31 $\pm$ 1.27
test 2	39	28	0.72	0.98 $\pm$ 0.01	−3.27	0.89	15.38	2.31 $\pm$ 1.28
test 3	37	26	0.70	0.98 $\pm$ 0.01	−3.16	0.84	15.10	2.09 $\pm$ 1.17
<i>period</i>								
Romano-British	6	6	1.00	1.00 $\pm$ 0.09	−1.79	1.00	n.a.	2.94 $\pm$ 1.63
Early Saxon	25	20	0.80	0.93 $\pm$ 0.01	−2.94	0.91	15.45	2.36 $\pm$ 1.19
Late Saxon	17	15	0.88	0.98 $\pm$ 0.02	−2.67	0.94	16.00	2.80 $\pm$ 1.41
<b>modern England</b>	<b>344</b>	<b>154</b>	<b>0.45</b>	<b>0.93 <math>\pm</math> 0.01</b>	<b>−4.06</b>	<b>0.70</b>	<b>13.59</b>	<b>1.84 <math>\pm</math> 0.93</b>
Cornwall	86	44	0.51	0.89 $\pm$ 0.03	−3.11	0.70	12.33	1.60 $\pm$ 0.83
Central and Northern	258	128	0.50	0.94 $\pm$ 0.01	−4.05	0.73	14.01	1.93 $\pm$ 0.97
<i>combined modern</i>	1156	379	0.33	0.94 $\pm$ 0.05	−4.65	0.65	13.92	2.03 $\pm$ 1.01

modern England (all Caucasians; for  $h$ ,  $p=0.046$ ), or for a range of probable source populations, including England, Denmark, Germany and Norway (for  $h$ ,  $p=0.009$ ; table 1). The relationship held after dividing modern England by region (table 1), even when the sample from Cornwall is excluded from the analysis (which is known to be relatively depauperate of variation; see Richards *et al.* 1996). A comparable diversity level to our ancient sample was seen for seventh and third century BC Etruscan samples ( $h=0.98$ ), but modern populations in southern Europe also have relatively higher levels of diversity than that in the north (Vernesi *et al.* 2004).

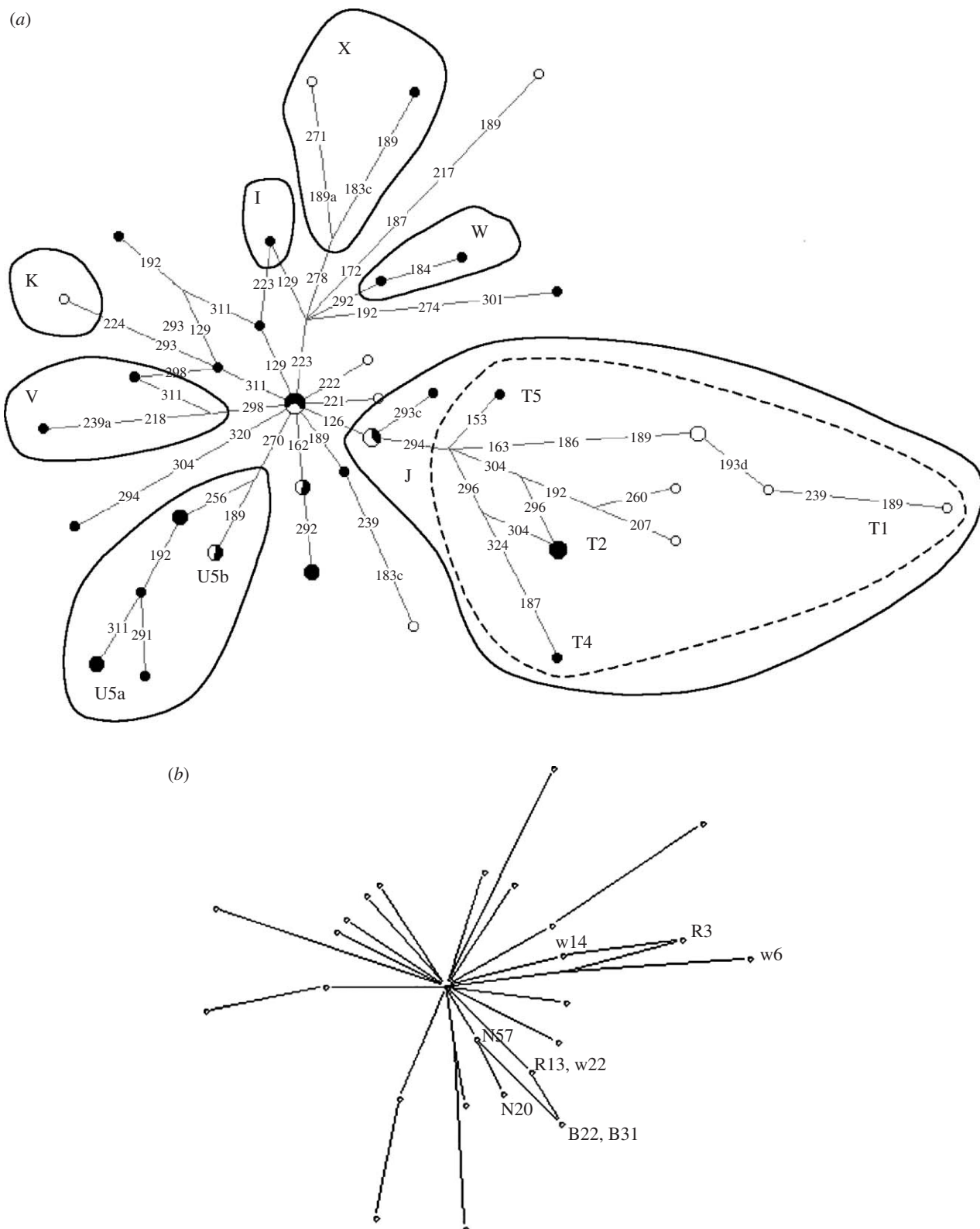
All archaeological sites were relatively short lived, spanning just over a century, indicating that the high level of diversity has not been inflated by extended temporal sampling. The difference in diversity level in our sample could be an artefact of small sample size, since most human haplotypes are rare, and therefore a small sample may not include many copies of the more common haplotypes (Helgason *et al.* 2000). To test this, we randomly resampled the full database for the modern England population for 10 subsamples of 48 individuals (where possible reflecting the geographical distribution of the ancient sample). They showed marginally inflated diversity as expected ( $h=0.897\text{--}0.965$ , mean=0.942), but all remained lower than  $h$  for the ancient sample. From a survey of 35 population samples from Europe and the Middle East ranging in size from 35 to 1199 (total  $N=6320$ ), only three populations: Turkey ( $N=72$ ); Palestine ( $N=117$ ); and Belarus ( $N=55$ ) showed levels of diversity as high as that seen in the ancient sample (data not shown).

Little networking was observed in the split decomposition tree (same structure found using Kimura 2P, HKY85 and uncorrected  $p$  genetic distance models; figure 1), and none showed parallel edges that might

indicate conflicting phylogenetic signals (see Holmes *et al.* 1999). The networking seen is most probably due to stochastic processes. All precautionary tests (see §2) resulted in measures of diversity higher than that seen for the modern populations (table 1). These very stringent tests show that diversity generated post-mortem or during amplification cannot fully explain the greater variation in the ancient sample.

Based on the assumption that modern England is more cosmopolitan, higher genetic diversity in the ancient sample was unexpected. One possible explanation is the stochastic loss of diversity. The Black Death in 1347–1351 resulted in the loss of a maximum of 50% of the population of Europe, although the impact was not uniform. Outbreaks reoccurred frequently until the fifteenth century (Ziegler 1969). During the second major impact, the Great Plague of 1665–1666, a fifth of the population in London died (Ziegler 1969). Because most human haplotypes are rare, lineages could have been forced to extinction even without a severe population bottleneck (millions remained). This is especially true if related families were lost when whole villages perished (either through shared susceptibility or environment). Furthermore, during the plagues some families were apparently more resistant than others (Clifford 1989), which could have led to high variance in the number of daughters surviving among different families and a consequent loss of mtDNA diversity at the population level. The fact that variance around the mean for  $h$  is greater in the ancient population compared with modern England is consistent with the loss of rare lineages (table 1; Bartlett's test,  $\chi^2_1=464$ ,  $p<0.0001$ ;  $\chi^2=46.8$ ,  $p<0.0001$  when a random subset of  $N=48$  from England is used).

Another possibility is strong selection for a mtDNA haplotype, suggested in our study by the observed frequency of the CRS (Cambridge reference



sequence) haplotype: 6.30% in the ancient sample and 19.53% ( $N=5529$ ) in modern Europe (21.70% in modern England,  $N=258$ ; average of 18.90% for eight modern European populations;  $N=32-54$ ). All of these showed significantly higher CRS frequencies compared with the ancient sample ( $\chi^2=4.72-6.21$ ;  $p=0.01-0.03$ ). Table 1 in Vernesi *et al.* (2004) also reveals a low frequency (7%) for the CRS in the

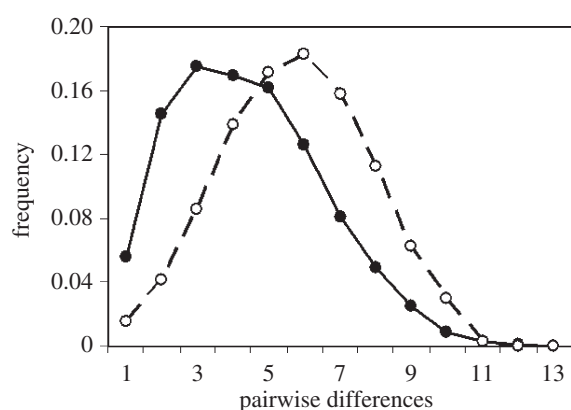


Figure 2. Mismatch distribution for the ancient DNA sample compared with that for modern England. Full line, Modern England; dashed line, ancient Britain.

gene diversity (by the most conservative estimate it would reduce  $h$  from 0.98 to 0.95, based on a proportional reduction of frequency equally distributed among the remaining haplotypes). Other studies have shown selection for human mtDNA lineages (e.g. Ruiz-Pensini *et al.* 2004), but our only evidence in this case is the differential frequency of the CRS haplotype, and this could also occur by drift (though all modern European populations have the high CRS frequency, not just England, and strong lineage sorting was not seen; see above). Helgason *et al.* (2003) provide evidence for drift distorting haplotype frequencies in the Icelandic population where the contribution of ancestral lines was apparently highly skewed. In reality, either drift or selection (or both) could have been involved, but further research will be required to estimate their relative importance (with modelling complicated by unavailable data, especially on the differential impact of disease on different families, and the extent to which Saxon villages were isolated).

We thank Jack Dumbacher, Alan Cooper, Trevor Anderson, Mandy Marlow, Lorraine Mephram, John Lucas, Liz Popescu, Peter Forster and Shane Richards.

- Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. 1995 Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753.
- Clifford, J. 1989 *Eyam plague, 1665–1666*. Sheffield, UK: The Print Centre.
- Comas, D., Calafell, F., Mateu, E., Pérez-Lezaun, A., Bosch, E. & Bertranpetit, J. 1997 Mitochondrial DNA variation and the origin of the Europeans. *Hum. Genet.* **99**, 443–449. (doi:10.1007/s004390050386)
- Gilbert, M. T. P., Hansen, A. J., Willerslev, E., Rudbeck, L. & Barnes, I. 2003 Characterization of genetic miscoding lesions caused by postmortem damage. *Am. J. Hum. Genet.* **72**, 48–61. (doi:10.1086/345379)
- Haak, W. *et al.* 2005 Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **310**, 1016–1018. (doi:10.1126/science.1118725)

- Helgason, A., Sigurðardóttir, S., Gulcher, J., Stefansson, K. & Ward, R. 2000 Sampling saturation and the European mtDNA pool: implications for detecting relationships among populations. In *Archaeogenetics: DNA and the population prehistory of Europe* (eds C. Renfrew & K. Boyle), pp. 285–294. Cambridge, UK: McDonald Institute for Archaeological Research, University of Cambridge.
- Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefansson, K., Ward, R. & Sykes, B. 2001 MtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am. J. Hum. Genet.* **68**, 723–737. (doi:10.1086/318785)
- Helgason, A., Hrafnkelsson, B., Gulcher, J. R., Ward, R. & Stefansson, K. 2003 A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosome. *Am. J. Hum. Genet.* **72**, 1370–1388. (doi:10.1086/375453)
- Holmes, E. C., Worobey, M. & Rambaut, A. 1999 Phylogenetic evidence for recombination in Dengue virus. *Mol. Biol. Evol.* **16**, 405–409.
- Huson, D. H. 1998 SPLITS TREE: a program for analysing and visualising evolutionary data. *Bioinformatics* **14**, 68–73. (doi:10.1093/bioinformatics/14.1.68)
- Magurran, A. E. 1988 *Ecological diversity and its measure*. Princeton, NJ: Princeton University Press.
- Marth, G. T., Czabarka, E., Murvai, J. & Sherry, S. T. 2004 The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* **166**, 351–372. (doi:10.1534/genetics.166.1.351)
- Miller, K. W. P., Dawson, J. L. & Hagelberg, E. 1996 A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *Int. J. Legal Med.* **109**, 107–113. (doi:10.1007/BF01369668)
- Petit, R. J., El Mousadik, A. & Pons, O. 1998 Identifying populations for conservation on the basis of genetic markers. *Conserv. Biol.* **12**, 844–855. (doi:10.1046/j.1523-1739.1998.96489.x)
- Piercy, R., Sullivan, K. M., Bensom, N. & Gill, P. 1993 The application of mtDNA typing to the study of white Caucasian genetic identification. *Int. J. Legal Med.* **106**, 85–90. (doi:10.1007/BF01225046)
- Richards, M. *et al.* 1996 Palaeolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**, 185–203.
- Ruiz-Pensini, E., Mishmar, D., Brandon, M., Procaccio, V. & Wallace, D. C. 2004 Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**, 223–226. (doi:10.1126/science.1088434)
- Thomas, M. G. *et al.* 2002 Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am. J. Hum. Genet.* **70**, 1411–1420. (doi:10.1086/340609)
- Töpf, A. L., Gilbert, M. T. P., Dumbacher, J. P. & Hoelzel, A. R. 2006 Tracing the phylogeography of human populations in Britain based on 4<sup>th</sup>–11<sup>th</sup> century mtDNA genotypes. *Mol. Biol. Evol.* **23**, 152–161. (doi:10.1093/molbev/msj013)
- Vernesi, C. *et al.* 2004 The Etruscans: a population genetic study. *Am. J. Hum. Genet.* **74**, 694–704. (doi:10.1086/383284)
- Ziegler, P. 1969 *The Black Death*. London, UK: Collins.